

MIT LIBRARIES DUPL



3 9080 00728848 0







HD28  
.M414  
no.

2511-  
89



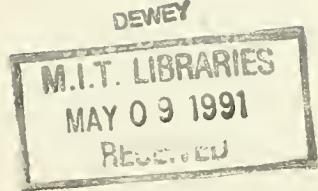
ANALYST FOLLOWING IN DIFFERENT  
INDUSTRY SECTORS

Ravi Bhushan and Patricia C. O'Brien

WP# 2511  
was 1911-89

January 1989

Comments are welcome. The authors are grateful to Robert Egan for valuable research assistance.







## Analyst Following in Different Industry Sectors

Ravi Bhushan and Patricia C. O'Brien

### 1. Introduction

Cross-sectional variation in information-related phenomena is often attributable to differences among industries. For example, cross-industry differences have been documented in the relative superiority of financial analysts over time-series models in forecasting earnings,<sup>1</sup> as well as in the marginal information content of earnings announcements by firms.<sup>2</sup> Our purpose in this paper is to explore cross-industry differences in information production in more detail. Specifically, we focus on one dimension of the problem: analyst following. Our objective is to identify economic factors associated with differences in analyst following, and to describe how these factors operate to generate observed differences in analyst following across industries.<sup>3</sup>

Following Bhushan (1988b), we view the number of analysts that cover a firm as a proxy for the level of economic resources devoted to private information acquisition by investors. Then, we expect supply decisions, that is, analysts' decisions to acquire and supply information, to be affected by industry factors that differentially affect the costs or

---

<sup>1</sup>Brown, Richardson, and Schwager (1987).

<sup>2</sup>Bhushan (1988a).

<sup>3</sup>Previous research [Bhushan (1988b), O'Brien (1987)] has indeed documented that some industries are more heavily followed by analysts than others. This is true even after adjusting for firm-specific factors like size and return variability.



benefits of collecting information. Examples of factors which may affect the costs of information collection, discussed further below, are industry stability and structure, and firm size. Factors which may affect the benefits from information collection are volatility and the extent of noise trading.

There may be differences in the demand for analysts' services, as well as the supply of those services. For example, since many institutional fund managers are held to standards of fiduciary responsibility, they may be called upon to substantiate the prudence of their investment decisions. One way they might do this is by supporting their decisions with expert judgement in the form of analysts' forecasts. Thus, the level of institutional holdings for a firm or an industry is likely to affect the demand for analysts' services. Similarly, factors which enter into investors' decisions, such as the growth potential of the industry or the systematic risk of firms in the industry, may affect demand for analysts' services. Below, we discuss more fully what factors we believe are likely to lead to industry differences in analyst following.

In the larger context of information production, our investigation can help lay a groundwork for providing economic explanations for cross-industry variation in other information-related phenomena. For example, if analyst following is a reasonable proxy for resources devoted to information production, we would expect to find lower levels of price response or information content for announcements from firms in more heavily-followed industries. We would likewise expect to observe a greater information-based advantage for financial analysts over forecasts based on time-series models in more heavily-followed industries.



Our approach in examining the economic determinants of cross-industry differences in information production is as follows. First, we demonstrate that cross-industry differences in analyst following appear in our sample. After controlling for various firm-specific characteristics which determine analyst following, a statistically significant amount of cross-industry variation remains. At this stage we make a digression to demonstrate and discuss a problem with the econometric specification of the model, namely that analysts' decisions to provide coverage about a firm and institutions' decisions to own shares in the firm are probably jointly endogenously determined. Resolution of this simultaneous-equations problem is still in progress. Finally, we separate the determinants of analyst following into firm-specific and industry-mean components, to examine how the effects operate. We are able to distinguish, by means of hypothesis tests on the estimated coefficients, between characteristics that operate only across industries but not within them, characteristics that operate across firms, independent of industry, and characteristics that operate across firms within industry.

The plan of the rest of the paper is as follows. In the next section we discuss various determinants of analyst following and our empirical proxies for them. Section 3 describes the data sources and some characteristics of the sample. In section 4 we present our results to date, and discuss several points which remain to be explored. In section 5 we present some brief concluding remarks.

## 2. Determinants of Analyst Following

In this section we discuss several factors, firm-specific or industry-



specific, that we expect to affect analyst following.<sup>4</sup> We view the number of analysts following a firm as a proxy for total resources spent on the information-gathering services, and examine factors that are likely to affect either the demand for or supply of these analyst services.

Specifically, we examine factors related to firm size and industry size, investor interest and perceptions, volatility and industry structure.

Other things equal, we expect analyst following to increase with firm size. If noise trading is heavier in larger firms, informed traders may be able to generate higher trading profits in larger firms because their information-based trading is better concealed. This increase in potential rewards would create a greater incentive for information-gathering for larger firms. In addition, if larger firms generally attract more investors, this could generate a larger volume of transactions business for analysts. The larger volume may implicitly lower analysts' costs of collecting information, causing a shift outward in the supply function for analyst services. Our primary proxy variable for size is the natural logarithm of the market value of equity. Several of the proxies mentioned below are size-related as well, and probably capture some size-related effects.

Investor interest in particular firms or industries is likely to affect the demand for analyst services. First, there is likely to be a straightforward demand from investors for information on which to base investment decisions. Second, for institutional investors and money managers there may also be a derived demand generated by standards of

---

<sup>4</sup>This discussion is based on Bhushan(1988b). The interested reader is referred to that paper for detailed discussion of the model.



fiduciary responsibility. If institutions are called upon to support the reasonableness of their decisions, analyst reports may help provide supporting documentation.

The relation between investor interest and analyst following is almost certainly not one-way. Analysts frequently perform or are associated with brokerage activities.<sup>5</sup> If analyst reports are used as part of the marketing of brokerage services to customers or potential customers, the effect may be to increase investor interest in followed firms. This argument suggests that analyst following and investor interest may be simultaneously determined. We explore this issue in section 4.

We use a variety of proxies for investor interest, including the number of institutions holding the stock, past earnings growth of the firm and the industry, the rating of the stock by Standard and Poor's Corporation, membership in the S&P 500, and the number of exchanges on which the firm's stock is traded. Obviously, several of these proxies have size-related characteristics as well, particularly the latter two. Other things equal, we expect greater analyst following to be associated with larger numbers of institutional investors, better past performance, higher ratings, membership in the S&P 500, and more exchange listings.

The systematic risk of the company and its industry may also affect analyst following through investor demand. Institutional investors may shun firms or industries with more systematic risk because adverse economic times

---

<sup>5</sup>This is particularly true for the database used in this study. I/B/E/S sells its forecast information to institutional investors. A careful distinction is made between buy-side analysts, the in-house analysts of institutions (potential customers), and sell-side analysts, those associated with brokers. Only the latter are included in the Summary forecast database.



are likely to have more impact on such firms or industries. Our argument here presupposes that institutions have an asymmetric loss function, and are more averse to negative outcomes than are other clienteles of investors. We can imagine this arising, for example, because the risk of lawsuits is greater for institutions who manage money in trust for others. This explanation indicates a negative relation between systematic risk and analyst following, *ceteris paribus*.

We expect that volatility will be positively related to analyst following. *Ceteris paribus*, an increase in return variability increases the probability that the expected return from private information could deviate by a large amount from the unconditional expected return. Thus, trading profits for informed investors are likely to be higher for more volatile firms or for firms in more volatile industries. This may increase demand for analyst services for such firms and industries.

If analysts face a fixed cost of becoming informed about a firm, then stability and industry structure may play a role in determining the analyst following. For example, industries with rapid turnover of firms because of frequent entry and/or failure may be more costly for analysts to follow than more stable industries. This would shift the supply curve for analyst services inward, creating a negative relation between the rate of entry/exit of firms in the industry and analyst following.

If there is some information common to many firms in an industry, then there may be economies of scale in collecting information. An example of the type of information which may be subject to scale economies is development of new technology or new products. We expect that, other things equal, an increase in the number of firms in the industry means more sources



of available information and thus a lower cost of information collection. This will result in a positive shift in the supply of analyst services and hence more analyst following.

In summary, in this section we have proposed several factors likely to affect analyst following and suggested directions for the effects. While we have not yet incorporated the variables related to industry stability and industry structure into our analysis, in future drafts we will consider the rate of entry/exit and the number of firms in the industry as explanatory variables. For some of the factors described above, it is difficult to say whether they act in firm-specific or industry-specific form or both. In section 4, we propose tests to examine whether a factor is firm-specific, industry-specific or both. In the next section, we describe our data sources and variables more explicitly.

### 3. Data Description

The analyst forecast data used here are from the I/B/E/S Summary tape produced by Lynch, Jones, & Ryan, covering the period from January 1976 through June 1988. All stock return data and SIC codes are from the Center for Research in Securities Prices (CRSP) 1987 Daily Stock and NASDAQ files. Stock ownership data are from Standard and Poor's (S&P Security Owners' Stock Guide).

The selection of the initial sample of firms and industries is described in Table 1. Firms are matched by CUSIP number from the I/B/E/S file to the CRSP files. A firm is omitted if it changed SIC codes between January 1976 and December 1986, or if it has no SIC code listed on CRSP. Of 5811 firms listed in the I/B/E/S file, 4254 are found in the combined CRSP



NYSE, AMEX, and NASDAQ files. These firms represent 318 3-digit SIC industries. We work only with industries in which there were at least five firms. This subset contains 155 industries, and 3887 firms.

Since several data items are hand-collected, we select approximately one quarter of the 155 industries for the final sample, to reduce the burden of data collection. We select every fourth SIC code in numerical order. Since SIC codes are numerically related (e.g., 3-digit codes are subsets of 2-digit codes), selecting every fourth industry gives us a broad-based set of industries. We are left with 38 industries, with 1104 firms.

Finally, we have so far collected data for the years 1985 through 1987. The number of firms in the 38 sample industries in these years is 678. The industry names and the number of firms in each industry in this final sample are listed in Table 2.

Since the number of analysts following any given firm tends to increase as the year end approaches, we collect the data near the fiscal year end, specifically in the eleventh month of the fiscal year. For example: for a December year-end company, we use the November I/B/E/S Summary list, and the November issue of the S&P Stock Guide. We obtain CRSP data from roughly the same time: defining day 0 to be the year-end date, we estimate parameters over trading days -244 to -45, and obtain the market value of equity at trading day -45.<sup>6</sup> This timing allows for a small time lag in the collection and publication of forecasts in the I/B/E/S Summary and S&P Stock Guide.<sup>7</sup>

<sup>6</sup> Forty-five trading days is approximately two calendar months. We require at least 50 return observations in the interval [-244, -45] to estimate parameters.

<sup>7</sup> The I/B/E/S Summary is produced monthly, in the third week of the month. Information in the database from which the Summary is produced is updated continuously as the information arrives at Lynch, Jones & Ryan. The



The variable names and their definitions are listed in Table 3. From the I/B/E/S file we obtain the number of analysts following the firm (#ANALYSTS), and a measure of the historical (5-year) growth in earnings per share (EPSGROW). From the CRSP Stock files we obtain systematic and non-systematic risk (BETA and RESIDSE) by estimating a market-model regression using the value-weighted Market Index,<sup>8</sup> and obtain the SIC code (SIC) and market value of equity (LNMKVAL). From the S&P Stock Guide we obtain information on the number of exchanges on which the firm's common stock is traded (#EXCHNG), the number of institutions holding the stock (#INSTNS), membership in the S&P 500 (SP500), and the S&P rating of the common stock.<sup>9</sup> We create a dummy variable (INVSTGRD) from the ratings, to indicate whether the stock is rated B or better, or below B.

Table 4 displays some characteristics of the sample by year. The first four columns contain the mean, standard deviation, median and interquartile range of the continuous variables. The last two columns contain an F-statistic and its associated p-value, or probability of Type I error, to test whether there is measurable variation across industries in the level of the variable. These comparisons are based on simple one-way analysis of

S&P Security Owners' Stock Guide states that the data in each monthly issue are "revised through the last business day of the prior month." Therefore, data from the November issue are current as of the end of October. This is approximately contemporaneous with the stock return data and the analyst data.

<sup>8</sup> In this draft, we have used the value-weighted index of NYSE and ASE stocks for the firms listed on CRSP, and the value-weighted index of NASDAQ stocks for OTC firms. In later drafts we will create a single value-weighted index to use for all stocks.

<sup>9</sup> This rating is similar to, but distinct from, the firm's bond rating. S&P rates the quality of the common stock on a scale of A to D, with + and - denoting variation within these categories.



variance, and do not control for any other variables. Notice that all variables, with the possible exception of the number of institutions holding the stock, exhibit some cross-industry variation.

#### 4. Empirical Tests of the Determinants of Analyst Following

The first step in our analysis is to verify that industry differences in analyst following exist in this sample, after controlling for firm-specific variables that may influence analysts' decisions to follow a firm. Consistent with earlier work, we find there is still variation in analyst following, after removing the effects of firm-specific variables. We discuss a complication in the model specification, namely that institutions' decisions to hold a particular firm are probably jointly endogenously determined with analysts' decisions to follow the firm. We demonstrate the nature of the problem of identifying the two endogenous variables separately. The solution to the problem is still in progress.

To examine the effects of firm-specific characteristics on analyst following, we perform the following regression:

$$\begin{aligned} \text{#ANALYSTS}_{it} = & a_{1t} SIC_i + a_{2t} SP500_i + a_{3t} INVSTGRD_i + b_{1t} \text{\#EXCHNG}_{it} \\ & + b_{2t} LNMKTVAL_{it} + b_{3t} \text{\#INSTNS}_{it} + b_{4t} \text{BETA}_{it} \\ & + b_{5t} EPSCROW_{it} + b_{6t} RESIDSE_{it} + e_{1it} \end{aligned} \quad (1)$$

separately for each year 1985 through 1987. That is, we regress analyst following on continuous variables proxying for breadth of market trading (#EXCHNG), size (LNMKTVAL), institutional interest (#INSTNS), risk (BETA), past performance (EPSCROW), and volatility (RESIDSE). The dummy variables for membership in the S&P 500 (SP500) and for investment quality (INVSTGRD)



are included as additional proxies for investor interest. The categorical industry variable is included to test whether, after controlling for these firm-specific characteristics that may affect performance, there is still variation across industries in analyst following.

We report the results of estimating equation (1) in Table 5. We show two specifications in each year, including and excluding the earnings performance variable, EPSCROW. This variable frequently had missing values, and so cut the sample size in each year quite dramatically, and caused two industries to be excluded entirely. Excluding EPSCROW allows us to use a larger set of observations in each year.

The results indicate that firm-specific characteristics explain significant variation in analyst following across years. Adjusted  $R^2$  values range between 73 and 80 percent. The coefficients on firm size and institutional following always have a strong positive association with analyst following, as expected.<sup>10</sup> The number of exchanges on which the stock is traded and the investment grade dummy vary in statistical significance from year to year, but the coefficients always have the expected positive signs. Systematic risk, past earnings performance and volatility do not add much explanatory power to the model. Membership in the S&P 500 is always statistically significant at the five percent level, and increases analyst following.

After controlling for these firm-specific factors, a significant amount of variation in analyst following across industries remains to be explained.

<sup>10</sup> These results on firm size are consistent with previous research, which suggests that more information is generated for larger firms. [E.g., Grant (1980), Atiase (1985), Collins, Kothari, and Rayburn (1987), Freeman (1987), and Bhushan (1988a)].



The F-statistic on the SIC categories is always significant, at Type I error levels of two percent or lower.

The statistical specification of equation (1), in terms of goodness-of-fit and residual plots, is quite good. However, two concerns about the economic specification are: (a) the possibility that the number of institutions holding a stock is jointly endogenous with the number of analysts following the stock, and (b) the possibility that the cross-sectional regression on the levels of these variables inflates the true covariation among the variables. Below, we explore these two issues and propose ways of addressing them.

To examine the number of institutions as a jointly endogenous variable, we first perform a "levels" regression using #INSTNS as the dependent variable, and the remaining variables as regressors:<sup>11</sup>

$$\begin{aligned} \text{#INSTNS}_{it} = & c_{1t} \text{SIC}_i + c_{2t} \text{SP500}_i + c_{3t} \text{INVSTGRD}_i + d_{1t} \text{\#EXCHNG}_{it} \\ & + d_{2t} \text{LNMKTVAL}_{it} + d_{3t} \text{\#ANALYSTS}_{it} + d_{4t} \text{BETA}_{it} \\ & + d_{5t} \text{EPSGROW}_{it} + d_{6t} \text{RESIDSE}_{it} + u_{lit} . \end{aligned} \quad (2)$$

The results of estimating equation (2) separately for each year, 1985-87, are in Table 6. The results here are similar to those from equation (1). Adjusted R<sup>2</sup> values are between 73 and 80 percent. Firm size and the number of analysts following the firm both have significantly positive coefficients. In this regression, the coefficient on the number of exchanges on which the firm's shares are traded always is significant, and systematic risk and volatility sometimes play a role. The coefficient on

---

<sup>11</sup> Badrinath, Gay, and Kale (1988) estimate and analyze a similar regression to explain institutional holdings.



BETA is negative, suggesting that, ceteris paribus, institutions prefer less systematic risk. Membership in the S&P 500 is strongly significant and increases institutional holding, though past earnings performance and investment grade are not significant.

If analysts' decisions to cover a firm and institutions' decisions to hold the stock are jointly endogenously determined, then the least-squares coefficients in equations (1) and (2) above are subject to simultaneous equations bias, and are inconsistent. To eliminate this bias, it is necessary to identify the system, either by specifying regressors which influence one but not the other endogenous variable, or by placing restrictions on the cross-equation covariances.

We found no compelling reason to restrict the equations on a priori grounds. That is, we felt that an argument could be made for each of the explanatory variables influencing both analysts' decisions to cover the firm, and institutions' decisions to hold the stock, and had no strong priors about cross-equation covariances. Instead, we will use statistical results from the above regressions and the differenced regressions below to identify the model. Since we are in the process of collecting data for earlier years, we will have an unused set of observations on which to test the simultaneous-equations formulation. Below, we describe what restrictions we feel are implied by the statistical results.

First, we explore the second of the concerns mentioned above about the economic specification of the regression models. Cross-sectional regressions on the levels of size-related variables may show only that there are size-related systematic differences among firms. If, however, equations (1) and (2) describe time-invariant structural associations among the



variables, then the relations should be maintained when the regressions are estimated in differenced form, where the differences are taken from year to year for a given firm. The differenced equations are:<sup>12</sup>

$$\begin{aligned}\Delta \# \text{ANALYSTS}_{it} = & b_{1t} \Delta \# \text{EXCHNG}_{it} + b_{2t} \Delta \text{LN MKTVAL}_{it} + b_{3t} \Delta \# \text{INSTNS}_{it} \\ & + b_{4t} \Delta \text{BETA}_{it} + b_{5t} \Delta \# \text{EPSGROW}_{it} \\ & + b_{6t} \Delta \text{RESIDSE}_{it} + e_{2it},\end{aligned}\quad (3)$$

and:

$$\begin{aligned}\Delta \# \text{INSTNS}_{it} = & d_{1t} \Delta \# \text{EXCHNG}_{it} + d_{2t} \Delta \text{LN MKTVAL}_{it} + d_{3t} \Delta \# \text{ANALYSTS}_{it} \\ & + d_{4t} \Delta \text{BETA}_{it} + d_{5t} \Delta \# \text{EPSGROW}_{it} \\ & + d_{6t} \Delta \text{RESIDSE}_{it} + u_{2it},\end{aligned}\quad (4)$$

where the symbol  $\Delta$  denotes year-to-year changes. We use both one- and two-year changes to estimate equations (3) and (4), and report the results in Tables 7 and 8, respectively.

If the phenomena we are measuring are true structural relations that do not vary over time, then the coefficients estimated in equations (3) and (4) should be the same as those estimated in equations (1) and (2), respectively. The results for equation (3), where the change in #ANALYSTS is regressed on firm-specific characteristics, are weaker and somewhat different from the results for equation (1), the levels regression. Only the number of institutions holding the stock remains consistently statistically significant, with coefficients similar to those reported in Table 5.

Four factors, other than a lack of structural relation, may contribute

<sup>12</sup>We have dropped the categorical and dummy variables from this specification. This is strictly correct for the SIC code variable, since we have excluded firms with industry changes from the sample. There are a few changes from year to year in S&P 500 membership and investment grade, but not enough to produce statistically significant results.



to the decline in statistical significance of many of the variables from the levels form to the differenced form. The first is that the parameters may shift over time. If this is the case, then there is no clear relation between the coefficients in the levels and differenced equations. A related possibility is that there are lead-lag relations among the variables which are not correctly specified in equations (1) and (2). If this is the case, then the differenced equations (3) and (4) are likewise mis-specified. A third possibility is that one- or two-year changes may not be sufficient to observe much change. For example, firms do not change or add exchange listings with great frequency, so  $\Delta\#EXCHNG$  may not have sufficient variation to generate measureable results. Finally, measurement error in any of the regressors is likely to play a greater proportional role in the differenced form of the equation than in the levels form. This would bias coefficients toward zero, perhaps to the extent that they are not measureably different from zero.

Equation (4), however, where the number of institutions holding the stock is regressed on firm characteristics, retains statistically significant coefficients for the number of analysts, size and volatility. Curiously, institutions seem more likely to hold more volatile stocks, other things equal, since the estimated coefficient on RESIDSE is positive.

The statistical results in Tables 7 and 8 suggest that size and volatility may be useful in identifying the pair of equations (3) and (4), since they appear important in the #INSTNS regression, but not in the #ANALYSTS regression. We are in the process of exploring this possibility.

For the remainder of the analysis in this paper, we drop #INSTNS as an explanatory variable in the regressions to explain analyst following, to



mitigate simultaneous equations bias problems. In Table 9 we report the results of the analyst regression, without #INSTNS as an explanatory variable. The results, as expected, show increases in explanatory power for variables linked to #INSTNS in the previous analyses. A comparison of Table 9 with Table 5 reveals that SP500, #EXCHNG and LNMKTVAL increase most in explanatory power when #INSTNS is omitted. The regression explanatory power drops slightly, as expected. The results in Table 9 are used below for comparison, as we attempt to distinguish between firm-level and industry-level determinants of analyst following.

Our attempt to distinguish firm-level from industry-level determinants is based on a comparison of between-industry and within-industry effects. For each characteristic, and for each firm and year in the sample we construct two variables. The first is the level of the characteristic for the firm itself, used in our previous regressions. The second variable is the industry mean for this characteristic. For example, for systematic risk,  $\text{BETA}_{it}$  refers to the estimated systematic risk of firm  $i$  in year  $t$ .  $\text{INDBETA}_{it}$  refers to the mean of betas of all firms in firm  $i$ 's industry, within sample.  $\text{INDBETA}_{it}$  captures variation across industries in the average level of systematic risk, presumably due to differences in risk associated with different lines of business. Other industry variables are defined and interpreted similarly.

We estimate the following regression:

$$\begin{aligned}
 \text{#ANALYSTS}_{it} = & g_{1t} \text{SP500}_i + g_{2t} \text{INDSP500}_i + g_{3t} \text{INVSTGRD}_{it} + g_{4t} \text{INDINVSTGRD}_{it} \\
 & + g_{5t} \text{#EXCHNG}_{it} + g_{6t} \text{IND#EXCHNG}_{it} + g_{7t} \text{LNMKTVAL}_{it} + g_{8t} \text{INDLNMKVAL}_{it} \\
 & + g_{9t} \text{BETA}_{it} + g_{10t} \text{INDBETA}_{it} + g_{11t} \text{EPSGROW}_{it} + g_{12t} \text{INDEPSGROW}_{it} \\
 & + g_{13t} \text{RESIDSE}_{it} + g_{14t} \text{INDRESIDSE}_{it} + e_{3it}
 \end{aligned} \tag{5}$$



separately for each year, 1985 through 1987. Notice that the industry dummy variables ( $SIC_i$ ) are dropped, since the vector of industry means for each variable is an exact linear combination of the firm-specific characteristic with the industry dummies. As above, we estimate (5) with two specifications, by excluding and including the earnings performance variables.

To understand the tests of results from equation (5), it is useful to consider an alternative, but equivalent, formulation. Instead of including the pair of regressors  $BETA_{it}$  and  $INDBETA_{it}$ , we could have included  $INDBETA_{it}$  along with the deviation of  $BETA_{it}$  from  $INDBETA_{it}$ . To facilitate discussion, we present below the portions of equation (5) pertaining to systematic risk under the two different formulations. For simplicity, we drop the year subscript  $t$  on both variables and coefficients.

$$\dots + g_9 BETA_i + g_{10} INDBETA_i + \dots \quad (6)$$

versus:

$$\dots + h_9 (BETA_i - INDBETA_i) + h_{10} INDBETA_i + \dots \quad (7)$$

From a comparison of (6) and (7), it should be clear that  $g_9 = h_9$ , and  $g_{10} = h_{10} - h_9$ . From (7), it should be clear that  $h_9$  measures the relation between analyst following and systematic risk within industry, while  $h_{10}$  measures the same relation between industries. We use these terms below to articulate our hypotheses.

If the effect of a particular characteristic results only from



variation across industries, and not from within-industry variation across firms, then the coefficient on the industry mean should be significantly different from zero, and the coefficient on the firm-specific portion should be indistinguishable from zero. In the example above, this corresponds to  $[g_9 = 0, g_{10} < 0]$ , or equivalently  $[h_9 = 0, h_{10} < 0]$ .<sup>13</sup> If both between- and within-industry variation influence analyst following, then in (6) both the coefficients will be significantly non-zero:  $[g_9 < 0, g_{10} < 0]$ . (In formulation (7),  $[h_9 < 0, h_{10} < h_9]$ ).

If the effect is purely firm-specific and industry plays no role, then in (6) the firm-specific variable would have a non-zero coefficient, while the industry-mean variable would have a zero coefficient,  $[g_9 < 0, g_{10} = 0]$ . (In (7),  $[h_9 < 0, h_9 = h_{10}]$ ). On the other hand, if there is no variation across industries in the effect of the characteristic, but all of its impact comes from within-industry variation, then in (7) the between-industry coefficient would be zero, while the within-industry coefficient would be measurably different from zero:  $[h_9 < 0, h_{10} = 0]$ . The equivalent test in formulation (6) is  $[g_9 < 0, g_{10} = -g_9]$ . Notice with these last two tests that we have a means of distinguishing firm-specific characteristics for which industry plays no role,  $[g_9 < 0, g_{10} = 0]$ , and characteristics which are firm-specific within each industry,  $[g_9 < 0, g_{10} = -g_9]$ . We return to this point below in the discussion of results.

The results from regression equation (5) are reported in Table 10. For most variables, the estimated coefficients and t-statistics for the firm-

---

<sup>13</sup>Our discussion above in section 2 suggested that the relation between systematic risk and the number of analysts should be negative. For variables expected to have a positive relation with the number of analysts, the signs on the hypothesis tests of coefficients should be reversed.



specific components are very similar to those in Table 9. The earnings performance variable (EPSGROW) does not contribute significant explanatory power, in either the firm-specific or the industry-mean component. However, the other variables all have results consistent with one of the interpretations listed above. We describe these interpretations in turn below.

Systematic risk (BETA), volatility (RESIDSE), and rating (INVSTGRD) appear to affect analyst following as between-industry effects, not as firm-specific effects. In other words, variation across industries in these characteristics determines analyst following to a statistically significant extent, while variation across firms (within industries or independent of industries) does not explain differences in analyst following. This result has some intuitive appeal, particularly in the case of systematic risk and volatility. These can quite plausibly be regarded as characteristics primarily related to a firm's line of business (therefore industry), and only secondarily to other firm-specific characteristics. A related interpretation is that, for these variables, the estimated industry means are better estimates of the underlying true determinant. For example, the industry average of estimated betas may provide a better estimate of systematic risk for each firm in the industry than the individual firm estimate.

Membership in the S&P 500 appears to determine analyst following within industries, but not between them. We can reject the hypothesis that  $g_1 = 0$  in each year, but cannot reject the hypothesis that  $g_1 = -g_2$ . Thus, industries with more S&P 500 firms are not more heavily covered by analysts, but within an industry, membership in the 500 adds about five analysts, on



average, to a firm's following.

Firm size and the number of exchanges on which the firm is traded appear to operate as firm-specific characteristics, independent of industry. That is, variation across industries in these two characteristics does not contribute significant explanatory power ( $g_6 = 0$ ,  $g_8 = 0$ ), but variation across firms does ( $g_5 > 0$ ,  $g_7 > 0$ ). The coefficient values indicate that analyst following increases by three or four when firm size doubles, and by one for every one or two new exchange listings.

In summary, with equation (5), we have examined the within-industry versus between-industry effects of several hypothesized determinants of analyst following. In this sample, we find that rating, systematic risk, and volatility influence analyst following at the industry level. Industries with higher S&P ratings of their common stock, with lower systematic risk, and with higher volatility have higher analyst following. Within an industry, membership in the S&P 500 increases analyst following. Firm size and the number of exchanges on which the firm's stock is listed influence analyst following at the firm level, independent of industry. Larger firms and firms listed on more exchanges have larger numbers of analysts following them.

## 5. Conclusions

In this draft of the paper, we have begun an analysis into the determinants of analyst following across industries. We have described a number of characteristics we expect will be related to analyst following, including measures of firm size, investor interest, volatility, and industry structure. Our tests to date indicate that it is possible to distinguish



some determinants of analyst following that operate at the firm level, independent of industry, and others that distinguish between and among industries.

We have also indicated directions for future exploration. Primary among these is the investigation of institutional ownership as a jointly-endogenous variable. We have provided statistical results that suggest a means of identifying the pair of equations for analysts and institutions, so that the two equations can be estimated simultaneously. Our intention is to use the observations included in this draft (1985 through 1987) to specify the model, and to collect data for earlier years to test the validity of the model.



## REFERENCES

Atiase, R., 1985, Predisclosure information, firm capitalization, and security price behavior around earnings announcements, *Journal of Accounting Research* 23, 21-36.

Badrinath, S.G., G.D. Gay, and J.R. Kale, 1988, Patterns of institutional investment and the managerial "safety-net" hypothesis, Working paper (Northeastern University and Georgia State University).

Bhushan, R., 1988a, Collection of information about publicly traded firms: theory and evidence, Working paper (M.I.T.), forthcoming in *Journal of Accounting and Economics*.

Bhushan, R., 1988b, Firm characteristics and analyst following, Working paper (M.I.T.), forthcoming in *Journal of Accounting and Economics*.

Brown, L., G. Richardson, and S. Schwager, 1987, An information interpretation of financial analyst superiority in forecasting earnings, *Journal of Accounting Research* 25, 49-67.

Collins, D.W., S.P. Kothari, and J.D. Rayburn, 1987, Firm size and information content of prices with respect to earnings, *Journal of Accounting and Economics* 9, 111-138.

Freeman, R., 1987, The association between accounting earnings and security returns for large and small firms, *Journal of Accounting and Economics* 9, 195-228.

Grant, E., 1980, Market implications of differential amounts of interim information, *Journal of Accounting Research* 18, 255-268.

O'Brien, P.C., 1987, Forecast accuracy of individual analysts in nine industries, M.I.T. Working paper #1940-87 (October).



TABLE 1

## Selection of Sample Firms and Industries

## Panel A: Intersection of Analyst and Stock Market Datasets

	<u>I/B/E/S</u>	<u>NYSE/AMEX</u>	<u>NASDAQ</u>
# of firms in file	5811	5019	7659
# matched by CUSIP <sup>1</sup>	4920	2545	2854
# omitted due to industry change	660	278	388
# retained in sample <sup>1</sup>	4254	2267	2466

## Panel B: Selection of Industries for Final Sample

	<u># firms</u>	<u># industries</u>
Intersection, from Panel A	4254	318
Industries with $\geq$ 5 firms	3887	155
Every 4th industry selected	1104	38
Data available 1985-87	678	38

1. The number of firms matched on I/B/E/S does not equal the sum of those matched on the NYSE/AMEX and those matched on NASDAQ, because some firms changed from OTC to one of the major exchanges (or the reverse) between 1976 and 1986, and so appear on both. We used CRSP's IPERM identification number to verify that these were identical firms.



TABLE 2

## SAMPLE INDUSTRIES

<u>3 - digit SIC</u>	<u>Industry Name</u>	<u># of sample firms</u>
131	Crude Petroleum and Natural Gas	46
201	Meat Products	6
205	Bakery Products	3
221	Broadwoven Fabric Mills, Cotton	6
232	Men's and Boys' Furnishings	11
243	Millwork, Plywood & Structural Members	5
264	Misc. Converted Paper Products	9
275	Commercial Printing	15
283	Drugs	50
287	Agricultural Chemicals	5
301	Tires and Inner Tubes	7
324	Cement, Hydraulic	3
332	Iron and Steel Foundries	6
343	Plumbing and Heating, Except Electric	6
349	Misc. Fabricated Metal Products	14
354	Metalworking Machinery	12
358	Refrigeration and Service Machinery	8
364	Electrical Lighting and Wiring Equipment	12
369	Misc. Electrical Equipment and Supplies	13
379	Misc. Transportation Equipment	6
384	Medical Instruments and Supplies	37
394	Toys and Sporting Goods	10
421	Trucking and Couriers Services, Except Air	15
483	Radio and Television Broadcasting	8
493	Combination Utility Services	32
503	Lumber and Construction Materials	3
512	Drugs, Proprietaries and Sundries	7
533	Variety Stores	5
566	Shoe Stores	4
591	Drug Stores and Proprietary Stores	10
612	Savings Institutions	29
621	Security Brokers and Dealers	13
633	Fire, Marine and Casualty Insurance	15
671	Holding Offices	225
721	Laundry, Cleaning and Garment Service	4
751	Automotive Rentals, No Drivers	4
805	Nursing and Personal Care Facilities	8
809	Health and Allied Services, NEC	6



TABLE 3

## Definitions of Variables

<u>Variable Name</u>	<u>Definition</u>
#ANALYSTS	The number of analysts listed in the I/B/E/S Summary file with estimates of current-year EPS.
EPSGROW	I/B/E/S' measure of the 5-year historical growth of EPS for the firm, "based on least square calculation of moving four quarters actual earnings for the latest 20 quarters."
#EXCHNG	The number of exchanges on which the stock is traded, as reported in the S&P <u>Security Owners' Stock Guide</u> .
#INSTNS	The number of institutions holding the stock, as reported in the S&P <u>Security Owners' Stock Guide</u> . Nearly 2700 institutions are covered, "including investment companies, banks, insurance companies, college endowments, and 13-F money managers." S&P obtains the data from Vickers Stock Research Corp.
INVSTGRD	Investment grade, based on the S&P rating of the stock. The variable takes the value 1 if the stock is rated B or better, and 0 otherwise. Unrated stocks are included among non-investment-grade.
SP500	Member of the S&P 500. The variable takes the value 1 if the firm is a member of the S&P 500, and 0 otherwise.
BETA	The estimated systematic risk, from a market model regression using the value-weighted index, over trading days -244 to -45. Day 0 is defined as the fiscal year end date.
RESIDSE	The residual standard error (x 100) from the market model regression described above in the definition of BETA.
LNMKVAL	The natural logarithm of the market value of equity (in \$1000s) on trading day -45.
SIC	The three-digit SIC code listed in CRSP for the firm.



TABLE 4  
Sample Characteristics, by Year

<u>Variable</u>	<u>Mean</u>	<u>Std. Dev.</u>	<u>Median</u>	<u>Q3-Q1</u>	<u>F(SIC)<sup>1</sup></u>	<u>p-value</u>
<b>1985:</b>						
#ANALYSTS	8.6	8.6	5	10	2.35	.0001
#INSTNS	80.7	118.7	37.5	76.5	1.66	.0100
#EXCHNG	2.4	1.7	1	3	4.75	.0001
LNMKTVAL	12.32	1.48	12.30	2.11	2.52	.0001
BETA	0.99	0.59	0.89	0.73	5.81	.0001
RESIDSE	1.93	0.87	1.71	0.95	4.02	.0001
EPSGROW	10.43	26.74	8.4	19.95	3.54	.0001
<b>1986:</b>						
#ANALYSTS	8.8	8.6	6	10	2.54	.0001
#INSTNS	90.9	128.4	45.5	81.8	1.99	.0006
#EXCHNG	2.3	1.7	1	2	3.81	.0001
LNMKTVAL	12.50	1.53	12.41	2.12	3.91	.0001
BETA	1.06	0.52	1.03	0.69	2.58	.0001
RESIDSE	2.15	0.99	1.91	1.07	5.01	.0001
EPSGROW	10.03	26.83	9.3	20.8	2.40	.0001
<b>1987:</b>						
#ANALYSTS	9.3	8.8	6	10	1.85	.0021
#INSTNS	111.78	153.28	53	103.5	1.30	.1158
#ESCHNG	2.5	1.8	2	3	3.04	.0001
LNMKTVAL	12.56	1.56	12.45	2.06	2.50	.0001
BETA	1.01	0.45	0.97	0.65	3.13	.0001
RESIDSE	2.26	0.95	2.06	0.98	2.85	.0001
EPSGROW	10.29	24.90	9.5	17.15	2.64	.0001

---

<sup>1</sup>The F-statistics test whether there are differences across industries in the levels of the variables, based on one-way analysis of variance. The p-values, or probabilities of Type I error associated with these tests, are reported in the following column.



TABLE 5

**Regression of Number of Analysts Following a Firm  
on Firm-Specific Characteristics, by Year**

**Coefficient Estimates (t-statistics in parentheses):**

<u>Regressor</u>	<u>1985</u>		<u>1986</u>		<u>1987</u>	
SP500	2.29 (2.54)	2.13 (2.97)	2.06 (2.29)	2.23 (2.98)	1.79 (2.09)	1.91 (2.57)
INVSTGRD	1.18 (1.51)	1.09 (2.36)	0.50 (0.80)	0.80 (1.78)	0.88 (1.39)	0.09 (0.20)
#EXCHNG	0.54 (2.32)	0.58 (3.40)	0.07 (0.32)	0.35 (2.14)	0.10 (0.53)	0.25 (1.59)
LNMKTVAL	2.30 (6.19)	1.72 (7.10)	2.45 (7.43)	1.93 (8.00)	2.32 (7.06)	2.04 (8.24)
#INSTNS	0.03 (6.52)	0.03 (10.93)	0.03 (7.88)	0.03 (10.71)	0.02 (8.56)	0.03 (10.74)
BETA	0.37 (0.55)	0.50 (1.20)	0.68 (1.10)	0.22 (0.56)	-0.08 (-0.11)	0.26 (0.52)
EPSGROW	-0.00 (-0.08)	---	-0.01 (-1.25)	---	-0.02 (-1.90)	---
RESIDSE	0.18 (0.32)	-0.09 (-0.28)	-0.18 (-0.41)	-0.18 (-0.70)	-0.42 (-1.07)	-0.23 (-0.87)

**F-statistics on Cross-Industry Differences in Intercept  
(p-values in parentheses):**

SIC	1.60 (.0208)	1.91 (.0013)	2.04 (.0007)	2.19 (.0001)	2.69 (.0001)	3.01 (.0001)
-----	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------

**Regression Summary Statistics**

Adj. R <sup>2</sup>	.739	.777	.777	.789	.796	.793
F	23.37	43.49	30.88	47.01	32.91	44.11
N	343	540	371	544	354	498



TABLE 6

**Regression of Number of Institutions Holding a Security  
on Firm-Specific Characteristics, by Year**

**Coefficient Estimates (t-statistics in parentheses):**

<u>Regressors</u>	<u>1985</u>		<u>1986</u>		<u>1987</u>	
SP500	62.31 (5.11)	68.40 (6.98)	79.61 (6.06)	79.42 (7.40)	77.83 (5.20)	84.66 (6.83)
INVSTGRD	-1.38 (-0.13)	-10.17 (-1.54)	-6.71 (-0.70)	-10.76 (-1.59)	-4.18 (-0.36)	3.05 (0.37)
#EXCHNG	6.86 (2.12)	7.34 (3.04)	9.40 (2.94)	8.09 (3.30)	7.68 (2.16)	7.23 (2.62)
LNMKTVAL	27.65 (5.25)	20.35 (5.82)	26.45 (5.05)	24.92 (6.80)	35.84 (5.88)	29.46 (6.71)
#ANALYSTS	4.92 (6.52)	6.27 (10.93)	6.09 (7.88)	6.47 (10.71)	7.93 (8.56)	7.83 (10.74)
BETA	0.50 (0.05)	-7.21 (-1.21)	-20.69 (-2.19)	-15.28 (-2.61)	-32.89 (-2.55)	-27.70 (-3.27)
EPSGROW	0.05 (0.32)	---	0.02 (0.13)	---	-0.14 (-0.74)	---
RESIDSE	-2.17 (-0.27)	2.24 (0.49)	4.28 (0.64)	4.97 (1.28)	17.45 (2.43)	15.91 (3.47)

**F-statistics on Cross-Industry Differences in Intercept  
(p-values in parentheses):**

SIC	1.31 (.1185)	1.87 (.0018)	2.35 (.0001)	2.52 (.0001)	1.75 (.0072)	2.08 (.0003)
-----	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------

**Regression Summary Statistics:**

Adj. R <sup>2</sup>	.738	.762	.779	.788	.793	.793
F	23.36	40.13	31.29	46.65	32.29	44.11
N	343	540	371	544	354	498



TABLE 7

**Regression of Changes in the Number of Analysts Following a Firm  
on Changes in Firm-Specific Characteristics, by Year**

**Coefficient Estimates (t-statistics in parentheses):**

<u>Regressor</u>	<u>1985-6</u>		<u>1986-7</u>		<u>1985-7</u>	
INTERCEPT	0.19 (1.28)	0.20 (1.83)	-0.51 (3.27)	-0.32 (-2.41)	-0.58 (-2.15)	-0.19 (-0.92)
Δ#EXCHNG	0.56 (1.93)	0.30 (1.37)	-0.13 (-0.65)	-0.09 (-0.48)	0.52 (1.98)	0.27 (1.25)
ΔLNMKTVAL	0.28 (0.80)	0.52 (2.32)	0.59 (1.79)	0.16 (0.62)	1.33 (2.85)	1.07 (3.32)
Δ#INSTNS	0.03 (5.50)	0.02 (5.76)	0.01 (3.53)	0.01 (2.79)	0.02 (3.99)	0.02 (4.39)
ΔBETA	-0.13 (-0.56)	-0.02 (-0.12)	-0.72 (-2.67)	-0.27 (-1.23)	-0.14 (-0.37)	0.09 (0.33)
ΔEPSGROW	0.01 (0.80)	---	-0.01 (-1.05)	---	0.00 (0.24)	---
ΔRESIDSE	0.17 (0.81)	-0.13 (-0.90)	0.21 (0.88)	0.16 (0.85)	0.15 (0.40)	-0.33 (-1.31)

**Regression Summary Statistics**

Adj. R <sup>2</sup>	.140	.119	.078	.020	.169	.127
F	8.72	12.93	5.13	2.53	9.17	11.45
N	293	452	307	437	248	369



TABLE 8

**Regression of Changes in the Number of Institutions Holding a Security  
on Changes in Firm-Specific Characteristics, by Year**

**Coefficient Estimates (t-statistics in parentheses):**

<u>Regressor</u>	<u>1985-6</u>		<u>1986-7</u>		<u>1985-7</u>	
INTERCEPT	8.64 (4.84)	8.05 (6.44)	19.22 (10.19)	14.53 (9.77)	20.56 (5.80)	18.46 (6.89)
Δ#EXCHNG	5.77 (1.62)	8.52 (3.22)	-4.03 (-1.42)	1.39 (0.60)	3.02 (0.81)	4.81 (1.60)
ΔLNMKTVAL	21.81 (5.42)	15.30 (5.87)	22.78 (5.19)	19.67 (6.31)	41.76 (6.84)	29.55 (6.94)
Δ#ANALYSTS	3.80 (5.50)	3.15 (5.76)	2.75 (3.53)	1.66 (2.79)	3.49 (3.99)	3.11 (4.39)
ΔBETA	2.08 (0.71)	2.52 (1.29)	-0.79 (-0.21)	0.08 (0.03)	-3.96 (-0.73)	-7.10 (-1.77)
ΔEPSGROW	0.05 (0.70)	---	0.12 (1.26)	---	0.04 (0.41)	---
ΔRESIDSE	4.27 (1.62)	2.45 (1.45)	9.56 (2.87)	7.41 (3.18)	22.59 (4.58)	14.25 (4.13)

**Regression Summary Statistics**

Adj. R <sup>2</sup>	.231	.190	.156	.102	.307	.211
F	15.41	22.40	10.24	11.02	18.61	20.39
N	293	452	307	437	248	369



TABLE 9

**Regression of Number of Analysts Following a Firm  
on Firm-Specific Characteristics,  
Excluding #INSTNS, by Year**

**Coefficient Estimates (t-statistics in parentheses):**

<u>Regressor</u>	<u>1985</u>		<u>1986</u>		<u>1987</u>	
SP500	4.39 (4.92)	5.25 (7.22)	4.94 (5.50)	5.57 (7.37)	4.50 (5.15)	5.12 (6.80)
INVSTGRD	1.31 (1.58)	0.97 (1.88)	0.39 (0.57)	0.60 (1.21)	0.97 (1.39)	0.22 (0.42)
#EXCHNG	0.81 (3.35)	1.00 (5.44)	0.37 (1.64)	0.72 (4.04)	0.37 (1.71)	0.56 (3.21)
LNMKTVAL	3.42 (9.76)	2.91 (12.15)	3.74 (12.00)	3.26 (14.24)	3.93 (13.19)	3.51 (15.31)
BETA	0.45 (0.62)	0.35 (0.76)	0.17 (0.25)	-0.27 (-0.62)	-1.09 (-1.39)	-0.58 (-1.07)
EPSGROW	0.00 (0.03)	---	-0.01 (-1.32)	---	-0.03 (-2.48)	---
RESIDSE	0.14 (0.23)	-0.03 (-0.08)	-0.08 (-0.17)	-0.05 (-0.17)	-0.01 (-0.02)	0.22 (0.76)

**F-statistics on Cross-Industry Differences in Intercept  
(p-values in parentheses):**

SIC	2.05 (.0007)	2.18 (.0001)	1.84 (.0035)	2.08 (.0003)	2.64 (.0001)	3.10 (.0001)
-----	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------

**Regression Summary Statistics**

Adj. R <sup>2</sup>	.704	.725	.736	.741	.749	.741
F	20.35	34.00	25.40	37.01	26.06	34.08
N	344	541	371	544	356	500



TABLE 10

**Regression of Number of Analysts Following a Firm  
on Both Firm-Specific and Industry-Specific Characteristics, by Year**

**Coefficient Estimates (t-statistics in parentheses):**

<u>Regressor</u>	<u>1985</u>		<u>1986</u>		<u>1987</u>	
SP500	4.27 (4.66)	5.25 (7.03)	5.06 (5.64)	5.57 (7.21)	4.48 (5.11)	5.12 (6.57)
INDSP500	-2.78 (-1.17)	-4.27 (-2.31)	-3.57 (-1.52)	-3.29 (-1.71)	-1.23 (-0.45)	-7.05 (-3.23)
INVSTGRD	0.93 (1.10)	0.97 (1.83)	0.40 (0.58)	0.61 (1.18)	0.86 (1.22)	0.22 (0.40)
INDINVSTGRD	3.75 (1.38)	2.38 (1.39)	7.41 (3.54)	3.88 (2.48)	5.29 (2.72)	5.32 (3.27)
#EXCHNG	0.84 (3.37)	1.00 (5.30)	0.36 (1.56)	0.72 (3.94)	0.36 (1.69)	0.56 (3.10)
IND#EXCHNG	-0.37 (-0.74)	-0.14 (-0.37)	0.47 (0.96)	0.20 (0.51)	0.74 (1.46)	0.71 (1.66)
LNMKTVAL	3.48 (9.68)	2.91 (11.83)	3.73 (12.07)	3.26 (13.91)	3.98 (13.31)	3.51 (14.79)
INDLNMKTVAL	2.39 (2.72)	1.74 (2.65)	0.29 (0.43)	0.12 (0.22)	-0.57 (-0.80)	-0.68 (-1.13)
BETA	0.28 (0.38)	0.35 (0.74)	0.41 (0.61)	-0.27 (-0.61)	-1.09 (-1.39)	-0.58 (-1.04)
INDBETA	-2.44 (-1.86)	-1.26 (-1.43)	-1.79 (-1.16)	-0.78 (-0.67)	-6.85 (-3.38)	-3.53 (-2.31)
EPSCROW	0.00 (0.09)	---	-0.02 (-1.37)	---	-0.03 (-2.46)	---
INDEPSCROW	0.01 (0.44)	---	0.02 (0.97)	---	0.04 (1.41)	---
RESIDSE	0.08 (0.12)	-0.03 (-0.08)	-0.14 (-0.30)	-0.05 (-0.16)	0.01 (0.03)	0.22 (0.73)
INDRESIDSE	4.18 (3.27)	3.19 (3.55)	3.82 (4.31)	2.47 (3.62)	7.14 (6.59)	4.49 (5.20)

**Regression Summary Statistics**

Adj. R <sup>2</sup>	0.680	0.708	0.728	0.727	0.741	0.720
F	53.05	110.88	71.66	122.18	73.40	108.73
N	344	541	371	544	356	500









Date Due

		Lib-26-67

MIT LIBRARIES



3 9080 00728848 0

